MOLECULAR AND CELLULAR BIOLOGY, Apr. 1987, p. 1465–1475
0270-7306/87/041465-11$02.00/0

Vol. 7, No. 4

Ingi, a 5.2-kb Dispersed Sequence Element from *Trypanosoma brucei* That Carries Half of a Smaller Mobile Element at Either End and Has Homology with Mammalian LINEs†

BRUCE E. KIMMEL,‡ ONESMO K. OLE-MOIYOI, AND JOHN R. YOUNG*
International Laboratory for Research on Animal Diseases, Nairobi, Kenya

Received 21 October 1986/Accepted 9 January 1987

A dispersed repetitive element named ingi, which is present in the genome of the protozoan parasite *Trypanosoma brucei*, is described. One complete 5.2-kilobase element and the ends of two others were sequenced. There were no direct or inverted terminal repeats. Rather, the ends consisted of two halves of a previously described 512-base-pair transposable element (G. Hasan, M. J. Turner, and J. S. Cordingley, Cell 37:333–341, 1984). Oligo(dA) tails and possible insertion site duplications suggested that ingi is a retroposon. The sequenced element appears to be a pseudogene copy of an original retroposon with one or more open reading frames occupying most of its length. Significant homologies of the encoded amino acid sequences with reverse transcriptases and mammalian long interpersed nuclear element sequences suggest a remote evolutionary origin for this kind of retroposon.

Eucaryotic genomes contain a variety of dispersed repetitive sequences. Many of these are mobile elements whose transposition to new sites in the genome accounts for their dispersion. For two major classes of mobile elements, the retroviruslike elements (16) and simpler retroposons (40), the transposition mechanism involves reverse transcription of an RNA intermediate. Retroviruslike elements are present in organisms as diverse as *Saccharomyces cerevisiae* (39), *Drosophila melanogaster* (41), and the mouse (19). They are characterized by long terminal direct repeats whose structure is intimately involved in the transposition process. They also encode reverse transcriptases which may be involved in their transposition (29, 45, 58).

A second class of elements, the retroposons, lack the terminal repeats of retroviruslike elements. They have been thoroughly reviewed by Rogers (40). The simplest of these are the processed pseudogenes, retrotranscribed copies of structural gene mRNAs, which may accumulate as small families of dispersed repeats. Other retroposons, while being transposed by a similar mechanism, have accumulated in much greater numbers. Some, for example, *Alu* elements, contain internal polymerase III promoters which may enable further rounds of retrotransposition to occur. These elements do not, however, appear to encode the enzymatic activities required for their retrotransposition.

Long interspersed nuclear elements (LINEs) are retroposons present in mammalian genomes in very large numbers (47, 48). They have characteristic A-rich or oligo(dA) tails at their 3′ ends and, like the retroviruslike elements, are flanked by short insertion site duplications. These two features are common to most retroposons and are probably direct results of the retrotransposition process, although it has not been proven that LINEs move by retrotransposition. Mouse and primate LINEs contain long open reading frames (ORFs), part of which encode amino acid sequences having significant homologies with reverse transcriptases from other sources (18, 25). Thus these elements may encode enzymes involved in their own transposition. The majority of LINE sequences are, however, truncated at their 5′ ends or contain disrupted reading frames. These LINE pseudogenes would be defective in terms of encoded enzymes and may be incapable of further retrotransposition. Thus only a small minority of the LINE sequences may be responsible for the generation of new family members (25). LINEs have been described only for mammals, although there is some similarity with *Drosophila melanogaster* F elements, which also have A-rich tails and short target site duplications (13). No retroposons of this kind have hitherto been described for protozoa.

In *Trypanosoma brucei*, a protozoan parasite, duplicative transposition of variable surface glycoprotein (VSG) genes plays a major role in the switching of VSG expression by which the organism evades the immune response of its mammalian host (4). There is no evidence to suggest that reverse transcription is involved in these transpositions. The existence of retrotransposition in this organism is, however, implied by the discovery of a dimer of a 512-base-pair (bp) ribosomal mobile element (RIME), which was found inserted into a rRNA gene in one *T. brucei* stock (17). RIME has an oligo(dA) tail and is flanked by short target site duplications, indicating that it is a retroposon.

In the course of studying repetitive sequences in the genome of *T. brucei*, we found that some RIME sequences were associated with a longer dispersed repetitive element. Structural analysis of this element suggests that it is a retroposon. Of the transposable elements that have been described for other organisms, the trypanosome element is most similar to mammalian LINE sequences.

MATERIALS AND METHODS

Trypanosomes. The trypanosome clones used in these studies were from isolates of the subgenus *Trypanozoon*. Details of the derivation of these cloned isolates have been described elsewhere: ILTaT 1.4 (32), MITat 1.2 (9, 31), ANTat 1.8 (27), and GUTat 3.1 (2) are *Trypanosoma brucei* subsp. *brucei*; ETat 1.5 (ETat 5 in reference 26) is *T. brucei*

* Corresponding author.
† ILRAD publication no. 488.
‡ Present address: Department of Molecular Biology, Wendel M. Stanley Hall, Berkeley, CA 94720.

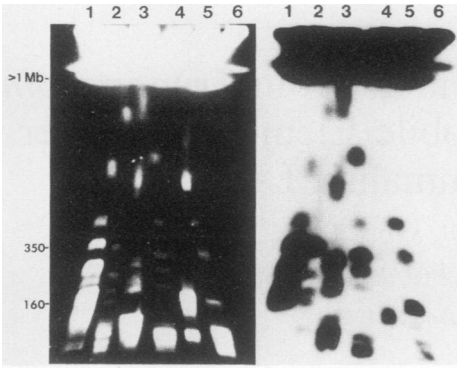

FIG. 1. Dispersed chromosome distribution of ingi elements. OFAGE of chromosome-sized DNAs from six cloned *Trypanozoon* isolates was carried out for 15 h at 300 V and a pulse interval of 30 s. The ethidium bromide-stained gel is shown on the left. DNAs: 1, ANTat 3.3 (*T. evansi*); 2, ETat 1.5 (*T. brucei* subsp. *rhodesiense*); 3, ANTat 1.8 (*T. brucei* subsp. *brucei*); 4, GUTat 3.1 (*T. brucei* subsp. *brucei*); 5, LITat 1.1 (*T. brucei* subsp. *gambiense*); 6, MITat 1.2 (*T. brucei* subsp. *brucei*). Sizes, indicated in kilobases unless marked otherwise, were measured by comparison with ladders of ligated lambda bacteriophage DNA. On the right is a Southern blot of the same gel hybridized with pgDR1.

subsp. *rhodesiense*; LITat 1.1 is *T. brucei* subsp. *gambiense*; and ANTat 3.3 is *T. evansi* (27).

DNA preparations. Trypanosome DNA was prepared as previously described (59). For orthogonal field alternation gel electrophoresis (OFAGE), purified trypanosomes (22) were embedded in 0.65% low-gel-temperature agarose and deproteinized as described by Van der Ploeg et al. (56). The OFAGE procedures described by Carle and Olson (7) were used.

Heteroduplex analysis. Plasmids pingi-1, pingi-2, and pingi-3 were digested with *Eco*RI and used to prepare heteroduplexes as described by Davis et al. (10). Photographs of heteroduplexes were analyzed by using a Science Accessories Corporation sonic digitizer interfaced to a microcomputer.

DNA probes. Radiolabeled DNA probes were prepared from plasmids by nick translation (28) and from M13 clones by extension of the 17-mer sequencing primer (Pharmacia, Uppsala, Sweden) in the presence of 100 μM each dATP, dGTP, and dTTP, 10 μM unlabeled dCTP, and 3 μCi of [α-^{32}P]dCTP (5,000 Ci/mmol; Amersham International plc., Amersham, United Kingdom) per μl, with the large fragment of *Escherichia coli* DNA polymerase I (Amersham International) in the buffer conditions recommended by the manufacturer. At least 2 kilobases (kb) of DNA complementary to the M13 clone DNA was copied under these conditions.

Copy number measurements. Approximate determinations of the copy numbers of segments of the ingi element were carried out by dot-blotting. Trypanosome DNA at 10 μg/ml in 10 mM Tris hydrochloride–0.1 mM EDTA (pH 8.0) was denatured by addition of 2 volumes of 0.5 M NaOH and incubation at 55°C for 10 min. Samples were stored on ice until 15 volumes of 1.2 M sodium acetate (pH 7.5) were added, and immediately aliquots of the solution were loaded onto replicate spots on a nitrocellulose filter in a Minifold I dot blot apparatus (Schleicher & Schuell, Keene, N.H.). Each dot was loaded with 100 ng of DNA. Standard samples containing lgB909 DNA and salmon sperm DNA to give a total DNA concentration of 10 μg/ml were loaded on the same filters in the same way to provide a standard scale. Filters were hybridized exactly as were Southern blots. After autoradiographic exposure, the amount of ^{32}P in each spot was determined by liquid scintillation counting of cut segments of the filters. The amount of probe sequence in the trypanosome DNA samples was estimated from a graph of the counts present in the lgB909 spots. Copy numbers were calculated on the basis of a total genome content of 0.1 pg of DNA (5).

Southern blots. Southern blots onto nitrocellulose filters were carried out as has been described previously (59), except that OFAGE blots were depurinated for 30 min in 0.2 M HCl at room temperature. Nitrocellulose filters were prehybridized in 4× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate–0.1% sodium pyrophosphate–750 μg of heparin per ml (49) for 4 to 8 h at 65°C. Hybridization conditions were identical to those for prehybridization, with the addition of 0.5×10^6 to 1.0×10^6 cpm of denatured radiolabeled probe per ml. Dextran sulfate at 10% (wt/vol) was added to the hybridization buffer for the OFAGE blot shown. After hybridization, all filters were washed twice in 0.1× SSC–0.1% sodium dodecyl sulfate at 65°C for 1 h.

DNA sequencing. All sequencing was performed by using the M13 dideoxy chain termination procedure (42). Subcloned fragments of the three ingi plasmids in pUC8 were used to generate shortened fragments by trimming with BAL 31 exonuclease. These were cloned into M13 vectors mp8, mp9, mp18, and mp19, producing clones from which overlapping sequences were obtained. Some sequences were also determined from cloned restriction fragments or fragments of sonicated DNA. The pingi sequences shown were determined over the entire length of both strands. More than 97% of the pgDR1 sequence was also determined on both strands. All included restriction enzyme sites were sequenced through. On average, each base was sequenced approximately six times.

RESULTS

Identification of a dispersed repetitive sequence. Repetitive sequences in the *T. brucei* genome were identified by hybridization of radiolabeled total DNA from ILTat 1.4(IL.467) to Southern blots of restriction enzyme digests of the same DNA. In *Sau*3AI digests, two major bands of hybridization, of approximately equal intensities, were detected and coincided with bands visible in ethidium bromide-stained gels. One of these bands was due to the 1.35-kb repeat sequence that encodes the leader sequence spliced onto the 5′ ends of *T. brucei* mRNAs (12, 15, 36). The other band, at 1.6 kb, was excised from a preparative gel and cloned into the *Bam*HI site of pUC8. Colonies were screened by hybridization with whole ILTat 1.4(IL.467) DNA, and the plasmid pgDR1 was isolated from a strongly hybridizing colony.

To determine the genomic distribution of sequences homologous to pgDR1, chromosomes from six cloned *Trypanozoon* isolates were resolved by OFAGE and blotted onto a nitrocellulose filter. The blot was probed with ^{32}P-labeled pgDR1. Most chromosomes from the six isolates contained sequences homologous to the pgDR1 probe (Fig. 1). The majority of the hybridizing sequences were on chromosomes larger than 10^6 bp. The intermediate-sized chromosomes (120 to 1,000 kbp) in most stocks contained at least one copy of the pgDR1 sequence, but the ratio of the hybridization signal to the ethidium bromide fluorescence showed substantial variation in the numbers of pgDR1 sequences carried in a given length of chromosome. These

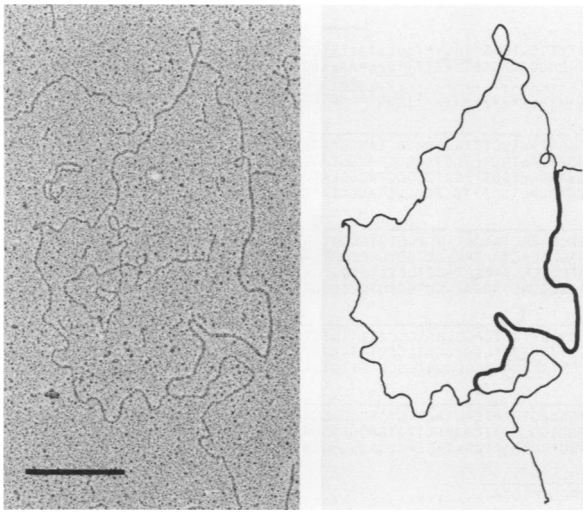

FIG. 2. Heteroduplex between ingi elements. The photograph on the left shows a heteroduplex formed between pingi-1 (circular DNA) and pingi-2 (linear). The scale bar represents a length of 0.5 μm. On the right is a line drawing of the heteroduplex photograph. The heavy and thin lines indicate double-stranded and single-stranded DNA, respectively.

differences were even more marked with the minichromosomes (50 to 120 kb). The experiment demonstrated that the pgDR1 sequence contained, or was part of, a dispersed repetitive element present in the genomes of these isolates of the subgenus *Trypanozoon*.

The reiteration of the pgDR1 *Sau*3AI fragment in the *T. brucei* genome could be explained either by the existence of a 1.6-kb tandemly repeated sequence or by the presence of a longer dispersed repeat of which it was an internal fragment. Southern blots with several other restriction enzymes showed a high degree of polymorphism in surrounding restriction enzyme sites and did not reveal strong hybridization to 1.6-kb fragments. This suggested that the pgDR1 sequence was part of a larger dispersed repeat rather than a unit from tandem arrays. Since the sequence was isolated from an African trypanosome, we chose the name ingi, the Kiswahili root adjective meaning many (20), for the dispersed repetitive sequence family identified with the pgDR1 probe. Under fairly stringent washing conditions, 0.2× SSC at 50°C, the sequence was not detected in cloned isolates of *T. congolense* or *T. vivax*, which belong to different subgenera.

Length and homology of cloned ingi elements. To characterize the ingi family, we sought clones containing whole elements. The clone lgB909 was obtained by screening a library of *Eco*RI total digest fragments of ILTat 1.4(IL.467) DNA in lambda gtWES with ^{32}P-labeled total ILTat 1.4(IL.467) DNA. The insert in this clone consisted of a 9.5-kb *Eco*RI fragment which contained a single 1.6-kb *Bam*HI fragment that hybridized with pgDR1. The lgB909 insert contained a group of restriction enzyme sites that had previously been observed in two genomic clones from a Charon 4A library of ILTat 1.3 DNA that contained different VSG genes (59). Hybridization experiments confirmed that these two clones carried pgDR1 sequences. The *Eco*RI fragments from the three lambda clones that carried pgDR1 sequences were subcloned into pUC8, producing the three plasmids pingi-1, pingi-2, and pingi-3. All three pairwise combinations of these plasmids were compared by heteroduplex analysis. A representative heteroduplex, formed from pingi-1 and pingi-2, is shown in Fig. 2. The double-stranded heteroduplex region is uninterrupted, showing that within the resolution of the method, there is continuous homology between the shared sequences over this region in these two clones. Similar results were obtained in the other comparisons. The lengths of the duplex region in 15 heteroduplexes from two plasmid combinations were measured as 5.37 ± 0.3 kb.

Physical maps of the three cloned ingi elements are shown in Fig. 3. The maps show the whole region of trypanosome DNA in each of the original lambda clones. They are aligned according to the homologous 5.4-kb ingi elements defined by the heteroduplex experiments (large open boxes). The VSG gene beyond the 3′ end of ingi-2 is the nontelomeric copy A of the ILTat 1.3 gene described by Young et al. (59). This gene is present on a very large chromosome that does not leave the slot in OFAGE gels. The 2.7-kb *Hin*dIII fragment

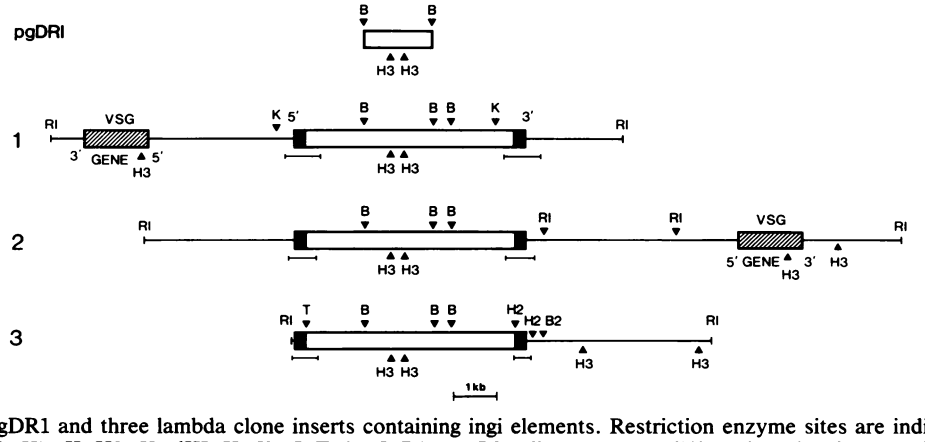

FIG. 3. Maps of pgDR1 and three lambda clone inserts containing ingi elements. Restriction enzyme sites are indicated as follows: B, *Bam*HI; B2, *Bgl*II; H2, *Hin*cII; H3, *Hin*dIII; K, *Kpn*I; T, *Taq*I; R1, *Eco*RI. All *Bam*HI, *Hin*dIII, and *Eco*RI sites are shown. Other enzyme sites are shown only where they were used in producing subclones for sequencing. Open and solid boxes show the regions of homology defined by heteroduplex analysis. The solid boxes at the ends of these regions are sequences homologous with RIME (see text). Shaded boxes are VSG genes. The VSG genes in ingi-2 is a nontelomeric copy of the ILTat 1.3 VSG gene. Small bars beneath the maps show the regions from which the sequence data shown in Fig. 4 were obtained. The 5′ and 3′ ends of the ingi elements are defined by the direction of the ORF in RIME (see below). The pingi plasmids are numbered according to the element numbering here. Ingi3 is from the lgB909 clone.

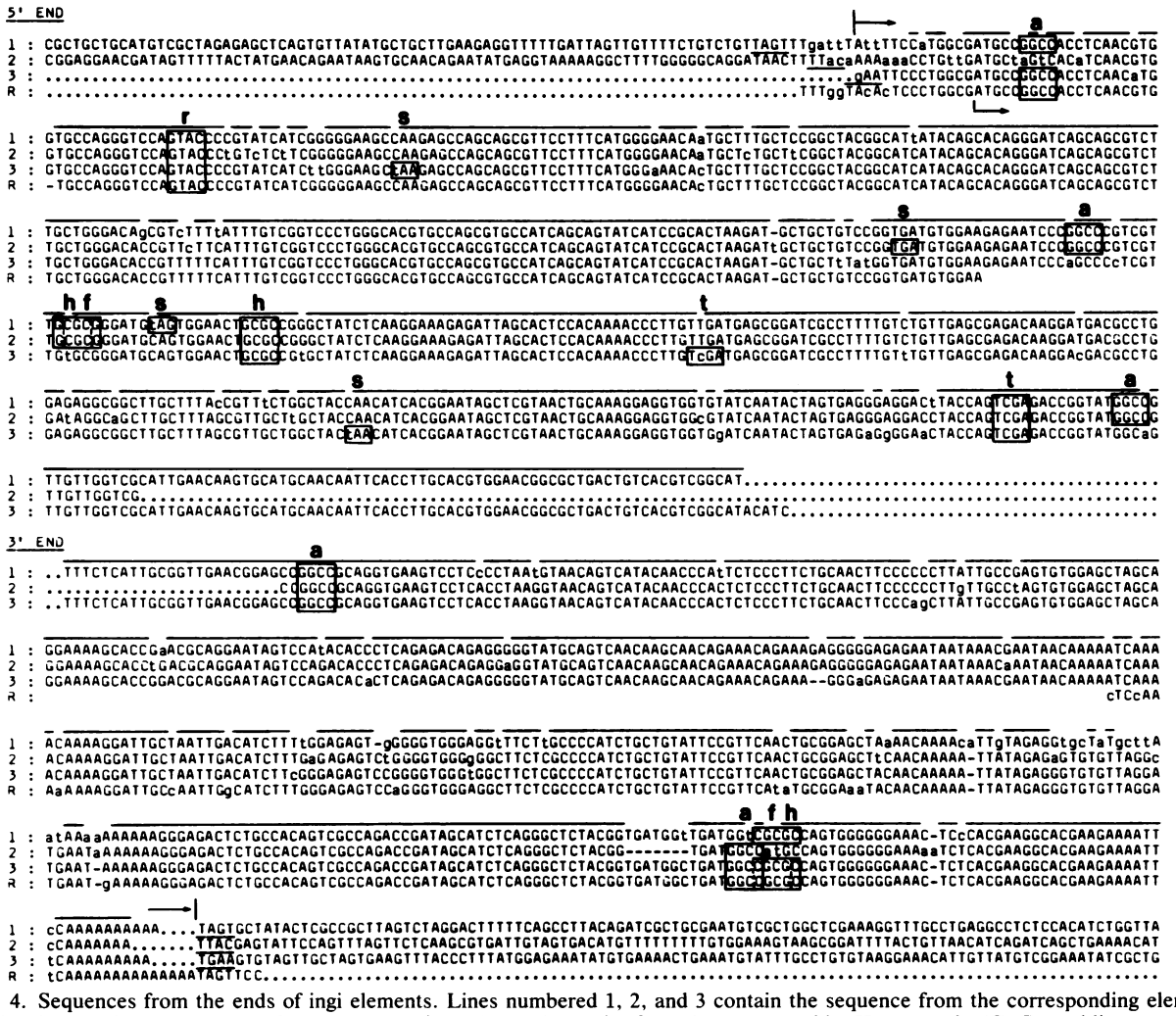

FIG. 4. Sequences from the ends of ingi elements. Lines numbered 1, 2, and 3 contain the sequence from the corresponding elements shown in Fig. 3. R is the sequence of one monomer from the tandem pair of RIMEs sequenced by Hasan et al. (17). Dotted lines represent the continuation of the DNA through the ingi elements. The RIME sequence is split into 5′ and 3′ halves and continues directly from one to the other. The precise limits of the RIME, including its oligo(dA) tail, are shown by vertical bars and arrows above the sequences. An arrow below the RIME sequence shows the ATG codon near the beginning of the RIME ORF. Hyphens represent gaps introduced into the sequences to obtain maximal alignment. Within the span of the ingi elements, a line is drawn above the sequences where all the sequences are identical. Bases that are not present in the majority of the sequences at any point are shown in lower case. Underlined bases at the ends of the ingi sequences are possible 4-bp insertion site duplications (see text). Stop codons in phase with the RIME ATG start codon in each of the 5′ ingi segments are boxed (boxes labeled s). Other boxes show the location of restriction enzyme sites: a, *Hae*III; f, *Fnu*DII; h, *Hha*I; r, *Rsa*I; t, *Taq*I.

beyond the 3′ end of ingi-3 also hybridized only to large chromosomes (results not shown). Thus at least two of these ingi elements came from large chromosomes. The presence of another VSG gene adjacent to ingi-1 was inferred from hybridization with ILTat 1.3 VSG cDNA and from heteroduplex analysis which showed that its orientation relative to the ingi element was opposite to that of the ILTat 1.3 gene (J. E. Donelson, unpublished data). Southern blots of restriction digests of the three ingi-containing plasmids probed with subcloned ingi fragments showed that further sequences homologous with ingi were absent from the cloned DNA segments. Thus these three elements were not parts of tandem arrays.

Definition of the ends of ingi. The sequences around the ends of the three ingi elements were determined to locate the ends precisely and to ascertain whether they contained direct or inverted repeats characteristic of certain classes of mobile elements. The location of the sequenced regions within the ingi-containing clones are shown by the small bars beneath the maps in Fig. 3. The sequences are compared in Fig. 4.

This analysis did not reveal either direct or inverted repeats at the ends of these ingi elements. Instead, it produced the completely unexpected finding that the extreme ends of the ingi elements were homologous with the two halves of the monomer of a previously described dimeric *T. brucei* mobile element, RIME (17). The sequence of one RIME (R) is included in Fig. 4. The 3′ limit of homology between the ingi sequences coincides exactly with the end of the RIME sequence. All three ingi elements end in an oligo(dA) stretch at this point, as does RIME. At the 5′ end, homology among all four sequences starts 6 bp after the start of the RIME sequence. Definition of the 5′ end of the ingi repeat from these data is therefore imprecise.

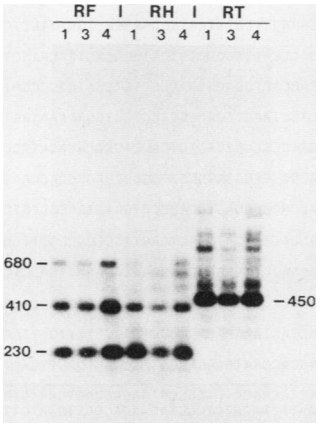

FIG. 5. Southern blot with ingi probes. Double digests of DNA from three trypanosome clones with *Rsa*I and either *Fnu*DII (RF), *Hha*I (RH), or *Taq*I (RT), resolved in a 1.9% agarose gel, were hybridized with an M13 probe containing the 342-bp 5′ end *Eco*RI/*Taq*I fragment from ingi-3. Sources of DNA: 1, ILTat 1.4; 3, ANTat 1.8; 4, MITat 1.2. Markers were fragments of φX174 DNA in the same gel.

Duplication of short (4 to 20 bp) segments of DNA at the insertion site is usually generated upon insertion of eucaryotic mobile elements whose transposition involves reverse transcription of an RNA intermediate (40). With ingi-1 and ingi-2, the 4 bp immediately following the oligo(dA) are found 1 (ingi-2) or 7 (ingi-1) bp farther 5′ than the RIME insertion site repeats identified by Hasan et al. (17). These short repeats may represent duplications resulting from the insertion of ingi elements. The ingi-3 sequence begins with an *Eco*RI site at one end of the cloned segment, but the underlined GAA may have been part of a repeat of the TGAA following the oligo(A) at the end of this element. If this were the case, the ingi-3 element, bounded by these repeats, would be 3 bp shorter than RIME at the 5′ end.

Over the regions of RIME homology, ingi-3 and RIME had the closest homology (96%). The degree of homology between the other pairs of sequences ranged from 90 to 93%. The sequence differences comprised 71 base differences and eight insertions or deletions. The 5′ end sequences were more conserved than those at the 3′ ends, particularly with respect to insertions or deletions. The divergence between the four sequences in Fig. 4 is substantially greater than that between the two RIME elements sequenced by Hasan et al. (17), which had 99.4% homology.

Ingi and RIME. The finding of RIME sequences at the ends of ingi elements raised the question of how many of these sequences were isolated RIME units and how many were part of ingi elements. Approximate copy number determinations were carried out by dot-blotting with pgDR1 and one M13 clone from either end of the ingi elements as probes. Using DNA from six *Trypanozoon* cloned isolates, we found that pgDR1 appeared to be present in about 200 copies (150 to 230 copies) per genome. A probe consisting of the 5′ 344 bp of ingi-3 (*Eco*RI/*Taq*I fragment; Fig. 4) gave similar results (145 to 220 copies). These values were similar to a previous estimate of 200 copies of RIME per genome (17). An M13 probe from the 3′ end of ingi-1 (*Kpn*I site to the end in Fig. 4) gave values between 225 and 415 copies per genome. On average, there were 1.5 times as many copies of the 3′ as of the 5′ end.

Since Hasan et al. (17) had suggested that most RIME sequences are present as isolated monomers and since the copy number experiments did not reveal the extent to which the RIME-homologous ingi ends were associated with the central pgDRI sequences, several Southern blot experiments were conducted to clarify the question of what proportion of RIME sequences might be present as part of ingi elements. One such experiment is shown in Fig. 5. The restriction enzyme sites involved in this experiment are shown boxed in Fig. 4. Interpretation of these experiments was complicated by the existence of restriction site polymorphisms in different ingi elements, by the lower retention of small fragments on nitrocellulose, and by different lengths of homology of detected fragments with the probe sequences. However, we generally observed substantially greater hybridization to fragments predicted from the ingi sequence (e.g., 230 bp in RF and RH digests and 450 bp in the RT digest in Fig. 5) than to those predicted from RIME monomers (410 bp in RF and RH digests in Fig. 5). Taking the above factors into consideration, we concluded from this and other experiments that although there are substantial numbers of both ingi elements and RIME monomers, there are more ingis than RIMEs.

Complete sequences of pgDR1 and ingi-3. Complete sequences were determined from the 1.6-kb *Sau*3AI fragment cloned in pgDR1 and the whole of the ingi element in pingi-3. The sequences are shown aligned in Fig. 6. The strand shown is that ending in an oligo(dA) tail in the ingi-3 sequence. Apart from a 3-bp insertion in the pgDR1 sequence relative to that of ingi-3 (star at position 2760 in Fig. 6), the alignment reveals 94.9% homology between the two sequences over a 1,560-bp region of ingi-3.

A stop codon map for the six possible reading frames of the complete ingi-3 sequence is shown in Fig. 7. The major feature of the map is a region of 3,360 nucleotides, starting at position 1036, which contains only a single stop codon. By analogy with other dispersed repetitive sequences, notably the mammalian LINEs, this may represent an ORF within a progenitor sequence of which ingi-3 is a pseudogene. Three stop codons within this region are present in the same frame in the pgDR1 sequence (boxed in Fig. 6), which may thus also be part of a pseudogene sequence.

On the same strand, the second reading frame contains only three stop codons in the region between positions 2 and 1160 which overlaps the longer ORF. Two of these, shown as dashed bars, are within the 5′ segment that was also sequenced in the other two ingi elements. These are shown boxed in Fig. 4 (boxes s). Both include bases which are unique to the ingi-3 sequence. This would be consistent with the presence of an ORF covering this region in an ancestral sequence from which the three ingi elements arose. The same frame has a further 689-bp ORF beginning at position 4253 which overlaps the 3′ end of the 3,360-bp ORF.

The putative ancestral ORF at the 5′ end of ingi-3 has an ATG codon at position 14 which coincides with the start codon in the ORF within RIME (17) (Fig. 4). The RIME reading frame, however, is shifted relative to the ingi frame after six amino acids by a single-base deletion. All three ingi sequences have stop codons in this reading frame within the region sequenced at the 5′ end. As was the case with ingi-3, the stop codons in ingi-1 and ingi-2 are the result of unique sequences in each element, a base change in ingi-1, and a single-base insertion in ingi-2 (Fig. 4). These unique sequence changes may be the result of accumulation of mutations in ingi pseudogenes.

Ingi ORFs are homologous with LINEs. The amino acid sequences encoded by the ingi ORFs were compared with

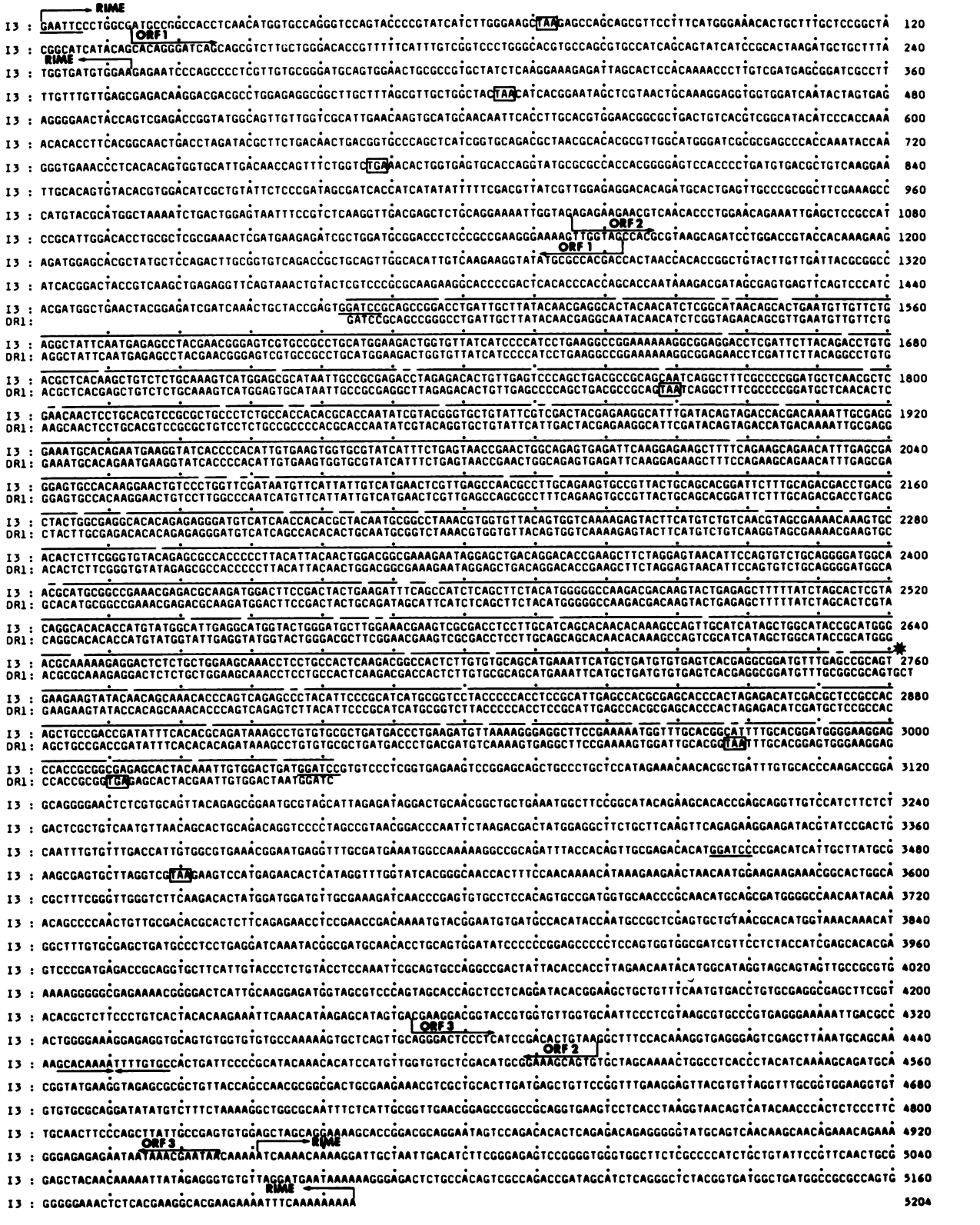

FIG. 6. Complete sequences of ingi-3 and the pgDR1 insert. I3 is the ingi-3 sequence. DR1 is the pgDR1 sequence. Identity between the two sequences is shown by overlining. The star at position 2760 in the ingi sequence shows the position of a 3-bp insertion in the pgDR1 sequence relative to that of ingi-3. The *Eco*RI site at the 5′ end of ingi-3, and *Bam*HI sites within it, are underlined. The regions with homology to RIME are delineated by arrows, as are three regions which are nearly open reading frames in the strand shown (see Fig. 7). Stop codons in phase with these frames in either sequence are boxed. A 16-bp palindrome in the ingi-3 sequence is shown by arrows beneath the sequence.

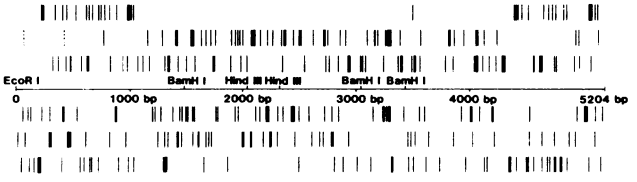

FIG. 7. Stop codon map of ingi-3. Stop codons in the six possible reading frames are shown by vertical bars. Those above the restriction map are from the strand shown in Fig. 6. The top frame starts with the GAA of the *Eco*RI site. The next two frames are successively shifted 1 base 3′. B-low the map, the three frames are mirrored so that the bottom frame corresponds to translation of the same codons as the top frame in reverse sense, etc. Two stop codons in the second 5′-3′ frame are shown as dotted bars because they are not present in the other two ingi elements sequenced.

the February 1985 version of the PIR protein sequence data base by using the DFASTP program (23). The highest-scoring homology was found with a hypothetical mitochondrial intron-encoded RNA maturase. Several authors have recently described homologies between these mitochondrion-encoded sequences and known or putative reverse transcriptases of retroviruses and eucaryotic mobile genetic elements (18, 25, 30, 34, 51, 52). Figure 8 shows an alignment of the relevant parts of the amino acid sequence encoded by the ingi-3 long ORF and pgDR1 with some of these sequences. Seven of the nine amino acids in this region that were found by Toh et al. (51) to be present in five viral polymerases are indicated. Seven of these are also present in one or the other of the ingi sequences. One of the two nonidentical positions is a favored substitution of Phe for Tyr in the sequence YXDD. This substitution is shared by two of the LINE sequences. The spacing of the homologous regions in the ingi sequences is more similar to that in the LINE elements than to those in the other sequences. These homologies suggest that the sequenced ingi element is a pseudogene of an original element which encoded an enzyme involved in nucleic acid metabolism, possibly a reverse transcriptase.

After addition to the PIR data base of the encoded sequence of the long ORF of a human LINE, L1Hs (18), the DFASTP program found a highly significant homology between ingi and the LINE sequence. The alignment of these sequences extended 5′ from the region of ingi included in Fig. 8. The RDF program (23) gave a score more than 20 standard deviations higher than the mean obtained with 50 shuffled L1Hs sequences. A difference of 10 standard deviations in this test is regarded as indicative of a significant evolutionary relationship (23). Thus this result provides evidence of an evolutionary relationship between the aligned regions of the ingi and L1Hs sequences. The extended alignment between these sequences is shown in Fig. 9, in which alignment with the mouse LINE L1MdA2 (25) is also included.

DISCUSSION

Ingi as a mobile element. The polydispersal of ingi sequences in the genome of *T. brucei* was shown by their chromosome distribution, their polymorphism in flanking restriction sites, and the isolation of the three cloned elements embedded within different sequences. Several lines of evidence suggest that ingi is a retroposon. As with other mobile elements which appear to transpose by way of an RNA intermediate, ingi elements terminate with 3′ oligo(dA) tails. Our sequence data did not allow the unequivocal definition of target site duplications, but possible 4-bp duplications were identified. If these were generated on insertion, the inserted elements must have been heterogeneous in length and sequence close to their 5′ ends, as has been shown for the 5′ ends of *D. melanogaster* F-element retroposons (13). Perhaps the most compelling evidence that ingi is transposable is that its ends are composed of the smaller element RIME, which was shown to have been transposed into a rRNA gene (17). The two elements may share a mechanism of transposition, although the transposed RIME dimer was flanked by 7-bp insertion site duplications.

While retroviruslike mobile elements have been found in genomes as diverse as those of mammals and yeasts, LINEs have been described only for mammalian genomes (40). A significant homology between ingi and LINE ORF-encoded sequences was found by an algorithm based on observed rates of amino acid changes in evolutionarily related proteins. Although this homology was much lower than that between LINEs from different mammals, it suggests that ingi and LINEs, or at least their regions of homology, are derived from a common ancestor sequence.

Ingi-3 contains three regions that may have been ORFs in a functional progenitor ingi sequence. As ingi-3 is probably a pseudogene, the two frame shifts separating these regions, as well as the four in-phase stop codons, may represent divergences from an original ingi element. The first frame shift occurs at a position similar to those in L1Md-A2 and

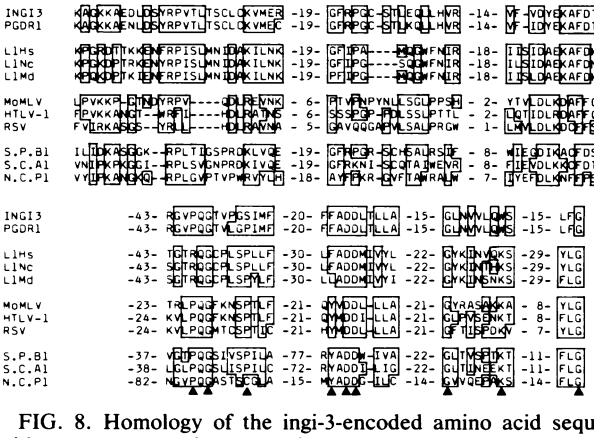

FIG. 8. Homology of the ingi-3-encoded amino acid sequence with reverse transcriptases and RNA maturases. The amino acid sequences and their sources are as follows. L1Hs and L1Nc are LINE sequences of human and prosimian origin (18). L1Md is the mouse LINE L1Md-A2 (25). MoMLV is from Moloney murine leukemia virus reverse transcriptase (46). HTLV-1 is human T-cell leukemia virus type 1 polymerase (44). RSV is Rous sarcoma virus polymerase (43). S.P.B1 is from the *Schizosaccharomyces pombe* cytochrome *b* intron maturase gene (21). S.C.A1 is from the *S. cerevisiae* cox1 gene a1 intron (3). N.C.P1 is from a *Neurospora crassa* mitochondrial plasmid (35). The ingi-3 and pgDR1 sequences are from this paper. The ingi sequences are aligned with the regions of the other sequences for which other authors have described homologies among LINE sequences, reverse transcriptases, and mitochondrial RNA splicing enzymes. Amino acid residues boxed are those for which at least six residues in the same position are identical or belong to the same group of amino acids related by evolutionary favored sustitutions as defined by Dayhoff et al. (11). The positions of 9 of the 10 amino acids that were found by Toh et al. (51) to be conserved in viral polymerases are indicated by triangles. Lengths of amino acid sequences separating the regions shown in the different genes are indicated between hyphens.

FIG. 9. Homology between encoded sequences of ingi-3 and mammalian LINEs. The LINE sequences are from mouse (L1Md-A2; [25]) and the consensus human sequence (L1Hs; [18]). The alignment shown is a composite of the alignment depicted in Fig. 8 (bottom line), the highly significant alignment found by the DFASTP program with the long ingi ORF, and a shorter alignment found by DFASTP with the shorter 5′ putative ancestral ingi ORF. The solid bar shows where the ingi reading frame switches. Symbols between the sequences respresent identical residues (*), residues belonging to the same group (:) where substitutions are favored in evolution (11), and other substitutions (·) with zero or positive scores in the 250 PAM matrix of Dayhoff et al. (11). Triangles below the sequences show the positions of the 10 amino acids conserved in viral polymerases (51). They are solid where the same amino acid identified in the polymerases is present in the ingi sequence (the half-shaded arrow is where the conserved P is present in pgDR1 but not ingi-3). Squares above the sequences are positions where conservative substitutions were in viral polymerases identified by Toh et al. (52). They are shaded if the corresponding amino acid in ingi-3 is also a conservative substitution in relation to those sequences.

Ty912, which may be skipped through during translation (8, 25). In relation to the positions of the amino acid-encoding homologies between the ingi and LINE sequences, however, the location of the first ingi frame shift is quite different. It interrupts the region of amino acid homology. It is therefore more likely that this shift is the result of an insertion mutation in this pseudogene. Thus the original ingi element may have had a single ORF occupying most of its length.

With pgDR1 and a 5′ ingi probe, the mean of estimated copy numbers in different trypanosome stocks was 200, or 1 per 400 kb if it is assumed that most pgDR1 sequences are within ingi elements. The overall density of ingi elements in the *T. brucei* genome is comparable with that in mammalian LINEs, 1 per 150 kb (48).

The 5′ long terminal repeats of proviruses and retrovirus-like elements contain promoters from which transcription necessary for transposition may occur. A 5′ array of tandem repeats in the L1Md-A2 sequence may serve the same function for this element (25). No analogous structure was found in the ingi sequences. Since no trypanosome promoter sequences have yet been identified, it is not possible to determine whether potential promoters, which would be necessary for retrotransposition, are present within the ingi sequence or in the 5′ flanking region of ingi-2 or ingi-3. We have obtained evidence from both runoff transcription and RNA slot-blot assays that ingi sequences are transcribed, but we have observed only heterogeneous transcripts in Northern blots. Thus it may be that ingi elements would have to rely upon fortuitous location near an external promoter if they are to retrotranspose. As is the case with LINE sequences (25), the number of ingi sequences, if any, which are capable of transposition is a matter of speculation, and the structure of any such elements may be different from that of the presumptive pseudogenes we have studied. Indeed, direct evidence for the retrotransposition of either LINEs or ingi elements is still lacking.

Ingi and RIME. It is clear that RIME and the sequences at the ends of ingi must have a common origin. RIME may have arisen by deletion of the 4.7-kb central region from an ingi element, or ingi may have arisen by insertion of the central sequences into a RIME. In either case, the boundaries of the insertion or deletion are identical in the three ingi elements we studied and in the two RIME sequences studied by Hasan et al. (17). Although these samples are small, this suggests that the two families of elements diverged after a single insertion or deletion event. The ingi elements are more divergent than the two RIMEs in the regions that can be compared. The extent of divergence of RIME from ingi is similar to the extent of divergence of the three ingi sequences. This suggests that the RIMEs may belong to a relatively recent sub-branch of ingi element evolution and thus may be of more recent origin than the ingi elements. This suggestion, however, relies on comparison of only two RIME sequences. Furthermore, there could be selective pressure to conserve the RIME ORF (17).

Deletion of the central portion of ingi, to produce the first RIME element, could have occurred either by splicing at the level of an RNA transcript or at the DNA level. However, there are no splicing signals at the appropriate points in the ingi sequence. Neither are there any direct repeats which might have mediated the excision of the central region by DNA recombination. Either of these features could, however, have been obscured by subsequent divergence of the sequences. There are other plausible mechanisms whereby RIME could have been generated from ingi elements, such as the insertion of a severely 5′-truncated ingi inside the 5′ end of a second element, but the available data do not allow assessment of these possibilities.

The alternative is that the original ingi element was derived from RIME. Many instances are known of retroposons inserting within each other (40). The point within the 3′ end of ingi at which the RIME homology

reappears (Fig. 4) is within a sequence very similar to the (NAx)y tails found at the 3′ ends of many retroposons (40). The (NAx)y sequence, and other A-rich regions within ingi, could be the relics of insertion events that were involved in the evolution of the original ingi element. However, part of the (NAx)y stretch was also present in RIME in our alignment of the sequences, which argues against its having resulted from an insertion into a RIME. We could not find any remnant of a target site duplication that might be expected if ingi arose by insertion of a shorter mobile element into RIME.

Whether an insertion or a deletion was responsible for the divergence of these two elements, analogies may be drawn between them and other mobile elements in which a shorter element, making up the ends of a longer one, carries the sequences required in *cis* for transposition and the longer element carries an encoded function required in *trans*. Thus, shortened *D. melanogaster* P elements are derivatives of the longer P factors which encode the transposase and may be mobilized by the presence of a complete P factor (37). The relationship between RIME and ingi may be of the same kind. RIME may be a disembowelled ingi element which retains terminal signals necessary for its transposition but depends upon the enzymatic activity encoded by a number of competent ingi elements. Alternatively, the combination of an original RIME with the internal sequences that encode enzymes required for its transposition might have produced a more efficient mobile element in the same way that some LINE sequences, embedded within shorter *Alu* retroposons, may rely upon the internal RNA polymerase III promoter of the enfolding element for their retrotransposition (33).

Does ingi have any important role in the trypanosome genome? The function, if any, of retroposons and retrovirus-like elements in eucaryotic genomes has been hard to define. They may represent essentially parasitic, or "selfish DNA" sequences, contributing nothing to the organisms whose genomes they occupy (14, 38). Their transposition does, however, have mutagenic effects, disrupting genes into which they are transposed or altering the expression of genes in their proximity, which may be of some advantage in the evolution of the host organism (40). The frequency of putative ingi retrotranspositions would clearly determine the extent to which any resulting mutagenesis could contribute to trypanosome evolution. Although extensive differences found in flanking restriction enzyme sites and chromosome distribution of ingi between different *T. brucei* stocks suggests a significant rate of transposition, we have found no differences between stocks known to be closely related. Thus it may be difficult to identify newly transposed ingi elements to confirm the occurrence of transposition and estimate its frequency.

Ty elements promote a variety of large-scale recombination events in the yeast genome, including chromosome translocations, duplications, and deletions (1, 39). Events of these kinds have been shown to occur, at least in the smaller chromosomes, in *T. brucei* (54) and may play significant roles in the process of antigenic variation and the evolution of VSG gene repertoires (53). It is thus possible that ingi elements promote recombination events that contribute to the evolution of VSG genes. Although two of the ingi elements we studied were in the neighborhood of nontelomeric VSG genes on large chromosomes, they were not present near other VSG genes we have studied (unpublished observations). Thus the association with VSG genes may not be significant.

Several abundant repetitive sequences in the *T. brucei* genome are arranged in tandem arrays (6, 12, 15, 24, 36, 50, 55). Thus, apart from the homologous sequences surrounding VSG genes (57), ingi and RIME sequences would appear to be the most abundant dispersed repetitive sequences in the *T. brucei* genome. It is possible that the trypanosome, living in intimate association with mammalian and insect host, acquired the original ingi sequence from one of these metazoan organisms. Unless this is the case, the finding of ingi in this protozoan suggests that this kind of retroposon may be more widely distributed among eucaryotic genomes than has been apparent (e.g., see reference 40). This would suggest a very early evolutionary beginning for these elements. The conserved sequences present in ingi, LINEs, retroviruslike elements, and the retroviruses themselves suggest that these may all be evolutionarily related. An early evolutionary origin of the retroposon sequences would then suggest that these were the precursors of the more complex elements. The question of whether the ingi element originated in a distant ancestor or was acquired from a host genome is therefore of importance in understanding the evolution of mobile elements that appear to use reverse transcription for their transposition.

ACKNOWLEDGMENTS

We are grateful to Paul Webster for DNA spreading and electron microscopy and to John E. Donelson for permission to refer to his unpublished heteroduplex experiments. We also thank A. J. Teale, S. P. Morzaria, and K. P. Iams for critical comments on the manuscript.

LITERATURE CITED

1. **Bernards, A., L. H. T. Van der Ploeg, W. C. Gibson, P. Leegwater, F. Eijgenraam, T. De Lange, P. Weijers, J. Calafat, and P. Borst.** 1986. Rapid changes of the repertoire of variant surface glycoprotein genes in trypanosomes by gene duplication and deletion. J. Mol. Biol. **190**:1–10.
2. **Black, S. J., R. S. Hewett, and C. N. Sendashonga.** 1982. *Trypanosoma brucei* variable surface antigen is released by degenerating parasites but not by actively dividing parasites. Parasite Immunol. **4**:233–244.
3. **Bonitz, S. G., G. Coruzzi, B. E. Thalenfeld, and A. Tzagoloff.** 1980. Assembly of the mitochondrial membrane system. J. Biol. Chem. **255**:11927–11941.
4. **Borst, P.** 1983. Antigenic variation in trypanosomes, p. 621–659. *In* J. A. Shapiro (ed.), Mobile genetic elements. Academic Press, Inc., New York.
5. **Borst, P., L. H. T. Van der Ploeg, J. F. M. Van Hock, J. Tas, and J. James.** 1982. On the DNA content and ploidy of trypanosomes. Mol. Biochem. Parasitol. **6**:13–23.
6. **Campbell, D. A., M. P. Van Bree, and J. C. Boothroyd.** 1984. The 5′ limit of transposition and upstream barren region of a trypanosome VSG gene: tandem 76 base-pair repeats flanking (TAA)90. Nucleic Acids Res. **12**:2759–2774.
7. **Carle, G. F., and M. V. Olson.** 1984. Separation of chromosomal DNA molecules from yeast by orthogonal-field-alternation gel electrophoresis. Nucleic Acids Res. **12**:5647–5664.
8. **Clare, J., and P. Farabaugh.** 1985. Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. Proc. Natl. Acad. Sci. USA **82**:2829–2833.
9. **Cross, G. A. M.** 1977. Antigenic variation in trypanosomes. Am. J. Trop. Med. Hyg. **26**(Suppl.):240.
10. **Davis, R. W., M. Simon, and N. Davidson.** 1971. Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. Methods Enzymol. **21**:413–428.
11. **Dayhoff, M. O., R. M. Schwartz, and B. C. Orcutt.** 1978. A model of evolutionary change in proteins, p. 345–352. *In* M. Dayhoff (ed.), Atlas of protein sequence and structure. National Biomedical Research Foundation, Silver Spring, Md.

12. **De Lange, T., A. Y. C. Liu, L. H. T. Van der Ploeg, P. Borst, M. C. Tromp, and J. H. Van Boom.** 1983. Tandem repetition of the 5′ mini-exon of variant surface glycoprotein genes: a multiple promoter for VSG gene transcription. Cell **34**:891–900.
13. **Di Nocera, P. P., M. E. Digan, and I. B. Dawid.** 1983. A family of oligoadenylate-terminated transposable sequences in *Drosophila melanogaster*. J. Mol. Biol. **168**:715–727.
14. **Doolittle, W. F., and C. Sapienza.** 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature (London) **284**:601–603.
15. **Dorfman, D. M., and J. E. Donelson.** 1984. Characterisation of the 1.35 kilobase repeat unit containing the conserved 35 nucleotides at the 5′-termini of variable surface glycoprotein mRNAs in *Trypanosoma brucei*. Nucleic Acids Res. **12**:4907–4920.
16. **Finnegan, D. J.** 1985. Transposable elements in eukaryotes. Int. Rev. Cytol. **93**:281–326.
17. **Hasan, G., M. J. Turner, and J. S. Cordingley.** 1984. Complete nucleotide sequence of an unusual mobile element from *Trypanosoma brucei*. Cell **37**:333–341.
18. **Hattori, M., S. Kuhara, O. Takenaka, and Y. Sakaki.** 1986. L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. Nature (London) **321**:625–628.
19. **Hawley, R. G., M. J. Shulman, and N. Hozumi.** 1984. Intracisternal A-particle genes: a family of copia-like transposable elements in *Mus musculus*. Cancer Cells **2**:273–279.
20. **Johnson, F.** 1939. A standard Swahili-English dictionary, p. 142. Oxford University Press, Nairobi, Kenya.
21. **Lang, B. F., F. Ahne, and L. Bonen.** 1985. The mitochondrial genome of the fission yeast *Schizosaccharomyces pombe*. The cytochrome b gene has an intron closely related to the first two introns in the *Saccharomyces cerevisiae* cox1 gene. J. Mol. Biol. **184**:353–366.
22. **Lanham, S., and D. G. Godfrey.** 1970. Isolation of salivarian trypanosomes from man and other mammals using DEAE cellulose. Exp. Parasitol. **28**:521–535.
23. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
24. **Liu, A. Y. C., L. H. T. Van der Ploeg, F. A. M. Rijsewijk, and P. Borst.** 1983. The transposition unit of variant surface glycoprotein gene 118 of *Trypanosoma brucei*. J. Mol. Biol. **167**:57–75.
25. **Loeb, D. D., R. W. Padgett, S. C. Hardies, W. R. Shehee, M. B. Comer, M. H. Edgell, and C. A. Hutchinson III.** 1986. The sequence of a large L1Md element reveals a tandemly repeated 5′ end and several features found in retrotransposons. Mol. Cell. Biol. **6**:168–182.
26. **Lumsden, W. H. R., and W. J. Herbert.** 1975. Pedigree of the Edinburgh *Trypanosoma* (*Trypanozoon*) antigenic types (ETat). Trans. R. Soc. Trop. Med. Hyg. **69**:205–208.
27. **Magnus, E., N. Van Meirvenne, and T. Vervoort.** 1981. Serological cross-reactivities of isotypic variable antigen type (iso-VATs) of the subgenus *Trypanozoon*. Ann. Soc. Belge Med. Trop. **62**:25–39.
28. **Meinkoth, J., and G. Wahl.** 1984. Hybridization of nucleic acids immobilized on solid supports. Anal. Biochem. **138**:161–174.
29. **Mellor, J., M. H. Malim, K. Gull, M. F. Tuite, S. McCready, T. Dibbayawan, S. M. Kingsman, and A. J. Kingsman.** 1985. Reverse transcriptase activity and Ty RNA are associated with virus-like particles in yeast. Nature (London) **318**:583–586.
30. **Michel, F., and B. F. Lang.** 1985. Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. Nature (London) **316**:641–643.
31. **Michels, P. A. M., A. Bernards, L. H. T. Van der Ploeg, and P. Borst.** 1982. Characterisation of the expression-linked gene copies of variant surface glycoprotein 118 in two independently isolated clones of *Trypanosoma brucei*. Nucleic Acids Res. **10**:2353–2366.
32. **Miller, E. N., and M. J. Turner.** 1981. Analysis of antigenic types appearing in first relapse populations of *Trypanosoma brucei*. Parasitology **82**:63–80.
33. **Miyake, T., K. Migita, and Y. Sakaki.** 1983. Some *Kpn* I family members are associated with the *Alu* family in the human genome. Nucleic Acids Res. **11**:6837–6846.
34. **Mount, S. M., and G. M. Rubin.** 1985. Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. Mol. Cell. Biol. **5**:1630–1638.
35. **Nargang, F. E., J. B. Bell, L. L. Stohl, and A. M. Lambowitz.** 1984. The DNA sequence and genetic organization of a *Neurospora* mitochondrial plasmid suggest a relationship to introns and mobile elements. Cell **38**:441–453.
36. **Nelson, R. G., M. Parsons, P. J. Barr, K. Stuart, M. Selkirk, and N. Agabian.** 1983. Sequences homologous to the variant antigen mRNA spliced leader are located in tandem repeats and variable orphons in *Trypanosoma brucei*. Cell **34**:901–909.
37. **O'Hare, K., and G. M. Rubin.** 1983. Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell **34**:25–35.
38. **Orgel, L. E., and F. H. C. Crick.** 1980. Selfish DNA: the ultimate parasite. Nature (London) **284**:604–607.
39. **Roeder, G. S., and G. R. Fink.** 1983. Transposable elements in yeast, p. 299–328. *In* J. A. Shapiro (ed.), Mobile genetic elements. Academic Press, Inc., New York.
40. **Rogers, J. H.** 1985. The origin and evolution of retroposons. Int. Rev. Cytol. **93**:187–279.
41. **Rubin, G. M.** 1983. Dispersed repetitive DNAs in *Drosophila*, p. 329–361. *In* J. A. Shapiro (ed.), Mobile genetic elements. Academic Press, Inc., New York.
42. **Sanger, F., A. R. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
43. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
44. **Seiki, M., S. Hattori, Y. Hirayama, and M. Yoshida.** 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA **80**:3618–3622.
45. **Shiba, T., and K. Saigo.** 1983. Retrovirus-like particles containing RNA homologous to the transposable element copia in *Drosophila melanogaster*. Nature (London) **302**:119–124.
46. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukaemia virus. Nature (London) **293**:543–548.
47. **Singer, M. F.** 1982. SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell **28**:433–434.
48. **Singer, M. F., and J. Showronski.** 1985. Making sense out of LINEs: long interspersed repeat sequences in mammalian genomes. Trends Biochem. Sci. **10**:119–122.
49. **Singh, L., and K. W. Jones.** 1984. The use of heparin as a simple cost-effective means of controlling background in nucleic acid hybridization procedures. Nucleic Acids Res. **12**:5627–5638.
50. **Sloof, P., J. L. Bos, A. F. J. M. Konings, H. H. Menke, P. Borst, W. E. Gutteridge, and W. Leon.** 1983. Characterization of satellite DNA in *Trypanosoma brucei* and *Trypanosoma cruzi*. J. Mol. Biol. **167**:1–21.
51. **Toh, H., H. Hayashida, and T. Miyata.** 1983. Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. Nature (London) **305**:827–829.
52. **Toh, H., R. Kikuno, H. Hayashida, T. Miyata, W. Kugimiya, S. Inouye, S. Yuki, and K. Saigo.** 1985. Close structural resemblance between putative polymerase of a *Drosophila* transposable genetic element 17.6 and *pol* gene product of Moloney murine leukemia virus. EMBO J. **4**:1267–1272.
53. **Van der Ploeg, L. H. T., and A. W. C. A. Cornelissen.** 1984. The contribution of chromosomal translocations to antigenic variation in *Trypanosoma brucei*. Philos. Trans. R. Soc. Lond. B **307**:13–26.
54. **Van der Ploeg, L. H. T., A. W. C. A. Cornelissen, P. A. Michels, and P. Borst.** 1984. Chromosome rearrangements in *Trypanosoma brucei*. Cell **39**:213–221.
55. **Van der Ploeg, L. H. T., A. Y. C. Liu, and P. Borst.** 1984.

Structure of the growing telomeres of trypanosomes. Cell **36**:459–468.

56. **Van der Ploeg, L. H. T., D. C. Schwartz, C. R. Cantor, and P. Borst.** 1984. Antigenic variation in *Trypanosoma brucei* analysed by electrophoretic separation of chromosome-sized DNA molecules. Cell **37**:77–84.
57. **Van der Ploeg, L. H. T., D. Valerio, T. De Lange, A. Bernards, P. Borst, and F. G. Grosveld.** 1982. An analysis of cosmid clones of nuclear DNA from *Trypanosoma brucei* shows that the genes for variant surface glycoproteins are clustered in the genome. Nucleic Acids Res. **10**:5905–5923.
58. **Wilson, S. H., and E. L. Kuff.** 1972. A novel DNA polymerase activity found in association with intracisternal A-type particles. Proc. Natl. Acad. Sci. USA **69**:1531–1536.
59. **Young, J. R., J. S. Shah, G. M. Matthyssens, and R. O. Williams.** 1983. Relationship between multiple copies of a *T. brucei* variable surface glycoprotein gene whose expression is not controlled by duplication. Cell **32**:1149–1159.